

## Grau en Matemàtiques

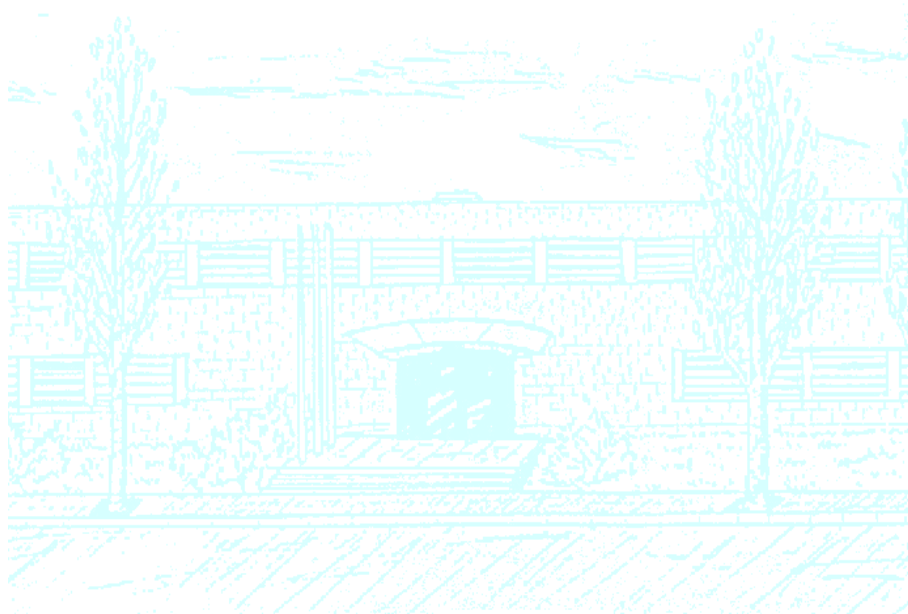
**Títol:** El Procés de Moran en grafs evolutius no dirigits

**Autor:** Paula Subías Beltrán

**Director:** Josep Díaz Cort

**Departament:** Llenguatges i Sistemes Informàtics

**Convocatòria:** 2013/2014





Universitat Politècnica de Catalunya  
Facultat de Matemàtiques i Estadística

Treball fi de grau

# **El Procés de Moran en grafs evolutius no dirigits**

Paula Subías Beltrán

Tutor: Josep Díaz Cort

Departament de Llenguatges i Sistemes Informàtics



Per tots aquells que han confiat en mi.



## Resum

**Paraules clau:** Grafs evolutius, Procés de Moran, Procés estocàstic, probabilitat de fixació, temps d'absorció

**MSC2000:** 60J10, 05C80

En aquest treball s'estudien alguns resultats sobre la modelització de l'evolució seguint el model d'infecció de Moran sobre grafs no dirigits. Concretament sobre un graf no dirigit, on els vèrtexs representen agents i les arestes les relacions entre aquests. Al principi, tots els agents estan sans. A temps 1, un agent és infectat. A partir d'aquest moment, es dona un procés estocàstic on l'agent infectat pot transmetre l'infecció a un veí *sa* amb probabilitat proporcional a un nombre real  $r > 0$  donat, que és anomenat adaptació genètica i representa l'habilitat de transmetre la infecció. De la mateixa manera, un agent *sa* pot curar a un vèrtex infectat amb adaptació genètica igual a 1. Aquest sistema és Markovià i té dos estats absorbents: fixació, l'estat on tots els vèrtexs han estat infectats, i extinció, on tots els vèrtexs han estat curats. Les principals quantitats estudiades per aquests sistemes són la probabilitat de fixació, probabilitat de què començant amb un únic vèrtex infectat tot el graf acabi infectat, i el temps d'absorció, el nombre esperat de passos que el sistema triga per arribar a un dels dos estats absorbents.

Aquest treball segueix part del material referenciat en [1,2,6]. Aquests articles són les fonts d'informació d'on hem extret els enunciats i demostracions d'aquest projecte. Per altra banda, l'autora d'aquest estudi és qui ha calculat les probabilitats de  $\mathbb{K}_n$  ( $r = 1$ ) i de l'estrella.

## Abstract

**Keywords:** Evolution on graphs, Moran Process, Stochastic Process, Fixation probability, Absorption time

**MSC2000:** 60J10, 05C80

In this project some recent results on the topic of modeling evolution of the Moran's infection model on undirected graphs are surveyed. Namely, given an undirected graph, where vertices represent agents and edges represent relationship between the vertices. At the beginning all agents are healed. At time 1, one agent gets infected. From then on, there is stochastic process where an infected agent can transmit the infection to a healthy neighbour with probability proportional to a given integer  $r > 0$ , which is called the fitness and represents the ability to transmit the infection. In the same manner a healthy agent can cure an infected node with a fitness of 1. This system is Markovian and it has two absorbing states: fixation, the state where all vertices end up being infected, and extinction, where all infected states have been healed. The main quantities studied for these systems are the fixation probability, probability that starting from a single infected vertex the whole graph ends infected, and the absorption time, the expected number of steps until the system stops in one of the two absorbing states.

This work follows part of the material in references [1,2,6], therefore the statement and proof for the theorems in the manuscript are taking from there. On the other hand, the author of this manuscript is responsible for the calculation of the fixations probabilities for  $\mathbb{K}_n$  ( $r = 1$ ) and the star.



# Índex

|  |           |
|--|-----------|
| <b>Introducció</b>   | <b>2</b>  |
| <b>1 Fonaments teòrics</b>   | <b>5</b>  |
| <b>2 Grafs evolutius</b>   | <b>9</b>  |
| 2.1 Procés de Moran . . . . .  | 11        |
| 2.2 Modelització estocàstica del procés de Moran en grafs evolutius                      | 15        |
| <b>3 Probabilitat de fixació</b>   | <b>18</b> |
| 3.1 Procés neutral de Moran . . . . .  | 18        |
| 3.2 Propietats de la probabilitat de fixació . . . . .                                   | 22        |
| <b>4 Procés de Moran per a grafs connexos no dirigits i <math>\Delta</math>-regulars</b> | <b>25</b> |
| <b>5 L'estrella</b>  | <b>32</b> |
| <b>6 Temps d'absorció</b>  | <b>38</b> |
| <b>7 Problemes oberts</b>  | <b>42</b> |
| 7.1 Doble estrella . . . . .   | 42        |
| 7.2 Camí . . . . .   | 43        |
| <b>Bibliografia</b>  | <b>46</b> |



# Introducció

Si observem el que passa al nostre voltant, ens sorprendrem de la quantitat d'interaccions que existeixen. Per exemple, les persones veïnes acostumen a obrir un compte de crèdit al mateix banc i a dur als seus fills al mateix col·legi, els directius adinerats tendeixen a freqüentar els mateixos llocs d'oci, els llibres d'escriptors reconeguts solen ser demanats per més llibreries que els escrits per aficionats,...

Des de la última dècada l'interès per les connexions entre agents que comparteixen un entorn ha anat en augment. La motivació d'aquest interès és entendre, entre d'altres, per què Internet ha crescut tan ràpidament en moltes zones o per què la informació, les epidèmies i la crisi financera es propaguen arreu del món amb una velocitat i una intensitat sorprenents. Aquests fenòmens que involucren xarxes<sup>1</sup>, incentius i comportaments de la gent; estan basats en les connexions entre les persones i en la manera en què les nostres decisions afecten al desenvolupament dels altres.

Cada xarxa té una configuració diferent, i poden estar formades per qualsevol tipus d'agent: persones, animals, gens, etc. Concretament, en aquest treball estudiarem la propagació de mutacions genètiques, molt relacionada amb la difusió d'epidèmies. Una malaltia pot contagiar ràpidament a tota una població o bé pot persistir durant molt de temps amb petits brots infecciosos, tal com s'explica al capítol 21 de [3]. Això dependrà no només del grau d'infecció de la malaltia sinó també del contacte amb gent sana que tinguin les persones infectades. És a dir, l'estructura de la població ens aporta molta informació sobre la propagació de l'epidèmia. Per tant, si som capaços de

---

<sup>1</sup>El terme xarxa es refereix a l'escenari on es reflecteixen les relacions del que volem estudiar.

modelar la xarxa subjacent, serem capaços d'entendre la difusió de la malaltia.

Per altra banda, el procés d'infeccions entre individus és diferent segons el context: propagació d'informació o d'idees, epidèmies o mutacions, etc. En el context de contagi social, per exemple quan es tracta d'adoptar una nova idea, les persones hem de prendre decisions que afecten a l'acceptació o no d'aquesta. En canvi, si parlem de malalties o mutacions, no hi ha cap presa de decisions. A més, al ser aquest procés tan complex i imprevisible, el més convenient és modelitzar-lo de manera aleatòria.

En biologia, hi ha diversos models que permeten definir l'evolució d'una població. En aquest treball, però, només tractarem un: el proposat per Patrick Moran i modificat, després, per Martin A. Nowak i altres científics. Patrick Moran va proposar un procés aleatori que descriu la dinàmica d'una població formada per dues espècies que competeixen per la supervivència. Aquest procés rep el nom de *procés de Moran*. En aquest model un individu és escollit per a reproduir-se i un altre per a morir. Després, l'individu mort és substituït per una còpia del primer. En el procés que Moran va proposar, ambdues seleccions es fan de manera aleatòria i les dues espècies tenen la mateixa *adaptació genètica*, és a dir, les dues espècies es reproduïxen amb idèntica probabilitat. L'evolució basada en aquest procés s'estudiava mitjançant equacions diferencials. Això és conseqüència de què només importava la variació del nombre d'individus de les espècies: no es considerava en el model l'evolució de cada individu en particular.

Aleshores, els biòlegs es van adonar de la importància que tenia establir connexions entre els individus. Com el procés original de Moran no les tenia en compte, el model podia considerar que un individu havia infectat a un altre quan, realment, no existia cap relació entre ells. Així, el model no era tan fidel a la realitat biològica com podia ser. Per aquest motiu, Lieberman i Nowak, tal i com podem veure a [6], van introduir una nova metodologia d'estudi: l'ús de grafs evolutius per representar les poblacions i, d'aquesta manera, establir relacions entre els individus, fet que el procés de Moran obviava.

A més, van decidir donar valor afegit al model tenint en compte que no totes les espècies es propaguen amb la mateixa velocitat: algunes, parlant en termes reproductius, són més avantatjoses que altres. Un clar exemple és con-

siderar els següents conjunts com a espècies: per una banda, les cèl·lules sanes del nostre cos i, per altra banda, les cèl·lules cancerígenes. D'aquestes dues espècies és ben conegut el fet de què les cèl·lules cancerígenes es reproduïxen més ràpidament que les normals<sup>2</sup>. Així, per tal de què els resultats s'ajustessin el màxim dins del possible a la realitat, van proposar afegir un nou factor: l'adaptació genètica de les espècies.

Aquest és un treball fonamentat en els progressos de Lieberman i Nowak que hem esmentat. Considerarem una població formada originalment per individus d'una mateixa espècie en la qual, esporàdicament, un dels seus membres pateix una modificació genètica: muta. D'aquesta manera, tractarem amb una població formada per dues espècies amb adaptació genètica diferent. A més, considerarem les relacions entre els agents. És a dir, estudiarem la propagació de mutacions genètiques mitjançant l'ús de grafs evolutius. La motivació és poder conèixer l'evolució d'una població d'individus entre els quals hi ha algun genèticament diferent a la resta.

---

<sup>2</sup>El lector pot trobar més informació sobre la dinàmica de les cèl·lules sanes que coexisteixen amb cèl·lules cancerígenes al capítol 12 de [7].

# Capítol 1

## Fonaments teòrics

Per poder endinsar-nos en l'estudi evolutiu de les mutacions donades en un conjunt d'individus, cal que parlem anteriorment sobre cadenes de Márkov. Si el lector volgues aprofundir més en aquests conceptes, recomanem la secció 6 de [4], font de la informació que a continuació presentarem.

Un procés estocàstic  $\mathcal{X} = \{X(t) : t \in T\}$  és un conjunt de variables aleatòries. L'índex  $t$  sovint representa el temps, i en aquest cas  $\mathcal{X}$  és una família de variables aleatòries  $X(t)$  que varien amb el temps.

Denotarem com  $X(t)$  l'*estat* del procés a temps  $t$ . A partir d'ara, utilitzarem  $X_t = X(t)$ . Si per tot  $t$ ,  $X_t$  pren valors d'un conjunt finit numerable, aleshores direm que  $\mathcal{X}$  és un procés d'espai discret. I si  $T$  és un conjunt infinit numerable direm que  $\mathcal{X}$  és un procés de temps discret.

A partir d'ara ens centrarem en un tipus especial de procés de temps discret<sup>1</sup>: el procés estocàstic d'espai discret  $X_0, X_1, \dots$  on el valor de  $X_t$  depèn del valor de  $X_{t-1}$  però no de la seqüència d'estats que s'han donat de temps 0 a temps  $t - 2$ .

**Definició 1.** Un procés estocàstic de temps discret  $X_0, X_1, \dots$  és una *cadena de Márkov* si

$$\begin{aligned} Pr(X_t = a_t \mid X_{t-1} = a_{t-1}, X_{t-2} = a_{t-2}, \dots, X_0 = a_0) \\ = Pr(X_t = a_t \mid X_{t-1} = a_{t-1}) = P_{a_{t-1}, a_t}. \end{aligned}$$

---

<sup>1</sup>Informem al lector que en aquest treball tractarem amb processos de temps finit.

Aquesta definició vol dir que l'estat  $X_t$  depèn de l'estat previ  $X_{t-1}$  però és independent de la història prèvia que explica com el procés ha arribat a  $X_{t-1}$ . Aquest fet rep el nom de  *propietat de Márkov*  i és això al que ens referirem quan diem que una cadena és de Márkov.<sup>2</sup> És important comentar que la propietat de Márkov no implica que  $X_t$  sigui independent de les variables aleatòries  $X_0, X_1, \dots, X_{t-2}$ ; només implica que el passat de  $X_t$  queda reflectit al valor de  $X_{t-1}$ .

Una cadena de Márkov pot ser representada mitjançant grafs dirigits on els vèrtexs representen els estats i les arestes les probabilitats d'anar d'un a l'altre. Així doncs, els vèrtexs representen les configuracions de  $X_t$  i les arestes dirigides entre nodes  $i$  i  $j$  tenen un pes que correspon a la probabilitat  $P_{i,j}$  d'anar de l'estat  $i$  al  $j$ . Fixem-nos que poden haver bucles de manera que durant un cert temps el procés romangui en un mateix estat. Els estats que no tenen sortides cap a altres estats s'anomenen *estats absorbents*. La resta, són anomenats *estats transitoris*. Notem que no totes les cadenes de Márkov tenen estats absorbents.

Sense pèrdua de generalitat podem assumir que l'espai discret d'estats d'una cadena de Márkov és  $\{0, 1, 2, \dots, n\}$  (o  $\{0, 1, 2, \dots\}$  si és un conjunt numerable infinit).<sup>3</sup> La probabilitat de transició

$$P_{i,j} = Pr(X_t = j | X_{t-1} = i)$$

és la probabilitat de què el procés vagi de  $i$  a  $j$  en un sol pas. La propietat de Márkov implica que la cadena de Márkov està definida de manera única per

---

<sup>2</sup>Es diuen cadenes de Márkov perquè una de les persones que va iniciar l'estudi dels processos estocàstics amb manca de memòria va ser Andrei Márkov, matemàtic rus del segle XIX.

<sup>3</sup>Informem al lector que en aquest treball, a partir de la següent secció, només tractarem amb cadenes de Márkov en espais discrets finits.

la *matriu de transició*

$$\mathbf{P} = \begin{pmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,j} & \cdots \\ P_{1,0} & P_{1,1} & \cdots & P_{1,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ P_{i,0} & P_{i,1} & \cdots & P_{i,j} & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

D'aquí tenim que per tot  $i$ ,  $\sum_{j \geq 0} P_{i,j} = 1$ . És a dir,  $\mathbf{P}$  és una matriu estocàstica.

Aquesta representació d'una cadena de Márkov mitjançant la seva matriu de transició és convenient per computar la distribució dels futurs estats del procés. Sigui  $p_i(t)$  la probabilitat de què el procés estigui a l'estat  $i$  a temps  $t$ . Sigui  $\bar{p} = (p_0(t), p_1(t), p_2(t), \dots)$  el vector que ens diu la distribució dels estats en una cadena a temps  $t$ . Sumant tots els possibles estats on podem estar a temps  $t - 1$  obtenim

$$p_i(t) = \sum_{j \geq 1} p_j(t-1)P_{j,i}$$

o

$$\bar{p}(t) = \bar{p}(t-1)\mathbf{P}.$$

Representem la distribució de probabilitats com un vector fila  $\bar{p}$  i observem que donada la distribució a temps  $t - 1$  podem obtenir la distribució a temps  $t$  aplicant la matriu  $\mathbf{P}$ .

Per qualsevol  $m > 0$ , definim la probabilitat de què la cadena vagi de l'estat  $i$  a l'estat  $j$  en exactament  $m$  passos com

$$P_{i,j}^m = Pr(X_{t+m} = j | X_t = i).$$

Si barregem aquest resultat amb la probabilitat de transició passant per l'estat  $i$  obtenim

$$P_{i,j}^m = \sum_{k \geq 0} P_{i,k} P_{k,j}^{m-1}. \quad (1.1)$$

Sigui  $\mathbf{P}^{(m)}$  la matriu amb valors equivalents a les probabilitats de transició



en  $m$  passos, per tant, a la fila  $i$  i columna  $j$  hi haurà  $P_{i,j}^m$ . A continuació, utilitzant l'equació (1.1) obtenim

$$\mathbf{P}^{(m)} = \mathbf{P}\mathbf{P}^{(m-1)}$$

i per inducció sobre  $m$ :

$$\begin{aligned}\mathbf{P}^{(m)} &= \mathbf{P}\mathbf{P}^{(m-1)} = \mathbf{P}\mathbf{P}\mathbf{P}^{(m-2)} = \mathbf{P}^2\mathbf{P}^{(m-2)} \\ &= \mathbf{P}^2\mathbf{P}\mathbf{P}^{(m-3)} = \mathbf{P}^3\mathbf{P}^{(m-3)} = \dots \\ &= \mathbf{P}^{m-k}\mathbf{P}^{(k)} = \mathbf{P}^{m-k}\mathbf{P}\mathbf{P}^{(k-1)} = \dots \\ &= \mathbf{P}^{m-2}\mathbf{P}^{(2)} = \mathbf{P}^{m-2}\mathbf{P}\mathbf{P} \\ &= \mathbf{P}^m.\end{aligned}$$

Per tant, per qualsevol  $t \geq 0$  i  $m \geq 1$ ,

$$\bar{p}(t+m) = \bar{p}(t) \mathbf{P}^m.$$

## Notació

- Donat un graf  $G$ , definirem  $G = (V, E)$  on  $V$  és el conjunt de vèrtexs i  $E$  el d'arestes.
- Sigui  $v \in V$ , aleshores definim  $\deg(v)$  com el grau del vèrtex  $v$ .
- El conjunt  $\mathbb{Z}$  és el conjunt dels enters.
- $t \in \mathbb{Z}$ , on  $t$  representa el temps.
- $\mathcal{P}(V)$  és el conjunt de particions de  $V$ .
- Un graf  $\Delta$ -regular és un graf on  $\forall v \in V \deg(v) = \Delta$ .

## Capítol 2

# Grafs evolutius

Els grafs evolutius ens permeten representar l'estructura d'un conjunt d'agents que comparteixen un entorn: els vèrtexs simbolitzen els agents i les arestes les relacions. Entre d'altres, com s'explica al capítol 1 de [3], els grafs evolutius poden descriure les relacions entre les persones de la xarxa social. Així, amb un cop d'ull podríem identificar quines són les persones amb més contactes dins de la xarxa. Si traslladem aquest exemple a un partit polític, podrem conèixer les persones amb més contactes que, en aquest context, podrien ser les més influents. Tanmateix, dins de l'àmbit social, la dinàmica del graf pot descriure l'evolució cultural i la propagació de nous invents i idees. L'èxit de la disseminació de quelcom nou depèn del grup de persones receptores. Per tenir una idea de com evolucionaria l'acceptació del que volem difondre, és convenient conèixer la influència entre les persones de la xarxa. On serà més fructífer començar la difusió? Als nuclis socials que, al graf, seran el conjunt de vèrtexs amb major grau. Per altra banda, podem conèixer, també, com és d'important una pàgina web a partir dels enllaços d'altres pàgines. Observem que, en aquest cas, tractaríem amb un graf dirigit ja que una pàgina pot fer referència a una altra sense la necessitat de què sigui recíproc. Com més pàgines referenciïn a una en concret, més important serà. Al graf estarà representada pel vèrtex amb més arestes incidents o sortints, depèn de la modelització que fem.

En aquest treball ens centrarem en l'evolució d'un conjunt d'individus entre

els quals hi ha algun que es diferencia genèticament de la resta. A aquest, l'anomenarem *mutant*. L'objectiu és conèixer la probabilitat i el temps esperat de què aquest mutant transformi genèticament a tota la població, és a dir, de què la població estigui formada únicament per individus mutants.

Als grafs evolutius amb els que tractarem, cada vèrtex representa un individu i les arestes determinen les relacions entre aquests. Els grafs poden ser dirigits o no dirigits depenent del procés evolutiu escollit (veure capítol 21 de [3] per a més informació). En aquest treball tractarem únicament amb grafs no dirigits seguint el procés evolutiu de Moran, el qual més endavant explicarem de manera més tècnica. Veiem un exemple per entendre quines són les relacions que descriuen les arestes d'aquest graf. Si hi ha una aresta entre el vèrtex  $i$  i el vèrtex  $j$  i un d'ells és mutant i l'altre no (suposem que l'individu representat pel vèrtex  $i$  és mutant i el representat pel vèrtex  $j$  no ho és) aleshores el mutant pot substituir amb una còpia seva a l'individu  $j$  i aquest últim pot substituir amb una còpia seva al primer. Això és degut a que s'influeixen mútuament. Aquest tipus de procés en epidemiologia s'anomena *SIS* (Susceptible Infectious Susceptible). Si féssim el símil amb un escenari social-cultural, informació i idees poden propagar-se entre els individus  $i$  i  $j$ .

Per estudiar l'evolució de poblacions que contenen individus mutants, cal introduir un concepte anomenat *adaptació genètica*<sup>1</sup> que denotarem com a  $r$ . L'adaptació genètica ens indica la capacitat d'un individu per a reproduir-se: a major adaptació, més gran és la probabilitat de què l'individu es reproduïxi. Considerarem que és una propietat d'un conjunt d'individus d'una espècie, no és pròpia únicament d'un individu. Veiem un exemple per entendre el seu significat. Suposem que tenim un conjunt d'individus de diferents espècies,  $A$ ,  $B$  i  $C$ , i que l'adaptació genètica de l'espècie  $A$  és major que la dels altres conjunts. Aleshores, els individus d' $A$  es reproduiran més ràpidament que els de  $B$  i  $C$ . És a dir, la selecció afavorirà a  $A$ .

Ara, assumirem que l'adaptació genètica de l'espècie sana és 1 mentre que la de l'espècie infectada és  $r \in \mathbb{R}^+$ . Aquesta assumpció és fa sempre i la gràcia serà estudiar la dinàmica del graf en tres casos: per  $r < 1$ ,  $r = 1$  i  $r > 1$ .

Aquest tipus de sistemes evolutius també han estat estudiats mitjançant

---

<sup>1</sup>Aquest terme rep el nom de *fitness* en anglès.

equacions diferencials i teoria de jocs, tal com s'explica al capítol 2 i al capítol 4 de [7], respectivament. L'avantatge de fer servir grafs és que es pot establir la influència entre individus particulars a través d'arestes, fet que simplifica el plantejament.

## 2.1 Procés de Moran

Considerem una població de mida  $n$ : actualment formada per individus de l'espècie  $A$  i de l'espècie  $B$ . A temps 0 aquesta població està constituïda exclusivament per individus de  $B$ . I així roman fins que un d'aquests pateix una modificació genètica que dóna lloc a l'espècie  $A$ , els mutants. L'adaptació genètica de l'individu  $A$  és  $r \in \Theta(1)$  mentre que la de  $B$  és 1. Si  $r < 1$ , la selecció afavorirà a  $B$ . Si  $r > 1$ , la selecció afavorirà a  $A$ . I si  $r = 1$ , la selecció serà neutral. Aquest resultat és conseqüència de la teoria de cadenes de Márkov, tal com podem veure al capítol 5 de [4].

El procés de Moran estableix que a cada pas un individu és escollit per a reproduir-se amb una probabilitat proporcional a la seva adaptació genètica. Aquest individu es duplicarà i la seva còpia reemplaçarà a un altre individu. Com ja hem esmentat anteriorment, seguint el model de Moran adaptat per Nowak (capítols 6 i 8 de [7]), l'elecció de l'individu que serà substituït no es fa sobre tota la població sinó que es realitza sobre el conjunt de veïns de l'individu seleccionat per a duplicar-se.

El fet de què a cada pas hi ha exactament un naixement i una mort assegura que la mida de la població roman sent  $n$ . Denotarem el nombre d'individus d' $A$  com a  $k$  i el de  $B$  com a  $n - k$ . El procés de Moran està definit en el conjunt  $k = 0, \dots, n$  (*espai d'estats*).

Per la teoria sobre cadenes de Márkov sabem que un cop el procés arribi a algun dels estats absorbents ( $|A| = 0$  i  $|A| = n$ ), romandrà allà per sempre. El procés es mou al conjunt d'estats transitoris per un període finit de temps. Finalment, la població estarà formada només per individus d' $A$  o de  $B$ . Malgrat que sigui un procés aleatori, un dels dos tipus reemplaçarà a l'altre.

Definim, doncs, l'algoritme que segueix el procés de Moran adaptat per Nowak.

Algoritme procés de Moran adaptat per Nowak.

Sigui  $G = (V, E)$  un graf no dirigit amb  $|V| = n$ . L'adaptació genètica dels mutants és  $r$  i la dels no mutants és 1. Comencem considerant que tots els vèrtexs estan habitats per individus no mutants.

- A  $t = 0$  es genera de manera aleatòria un individu mutant a  $V$ .
- Per  $t > 0$  assumim que hi ha  $k$  mutants i  $(n - k)$  no mutants. Definim l'adaptació genètica total com  $W_t = k r + (n - k)$ .

**mentre que** l'estat no sigui un dels estats absorbents

**escollim**  $u \in V$

**si**  $u$  és mutant l'elecció té probabilitat  $\frac{r}{W_t}$

**altrament** l'elecció té probabilitat  $\frac{1}{W_t}$

**fi si**

**escollim**  $v \in V$  tal que  $(u, v) \in E$  amb probabilitat  $\frac{1}{\deg(u)}$

**si**  $v$  és mutant

**si**  $u$  és mutant **aleshores** tornem a escollir  $u$ .

**altrament**  $u$  no és mutant **aleshores** una còpia de  $v$  substitueix a  $u$ : ara  $u$  és mutant.

**fi si**

**altrament**  $v$  no és mutant

**si**  $u$  és mutant **aleshores** una còpia de  $v$  substitueix a  $u$ :  $u$  ja no és mutant.

**altrament**  $u$  no és mutant **aleshores** reescollim  $u$ .

**fi si**

**fi si**

**fi escollim**

**fi escollim**

**fi mentre**

Notem que a cada pas el nombre de mutants evoluciona en  $\{-1, 0, 1\}$ .

Com aquest procés estocàstic té dos estats absorbents, que són l'estat en què tots els individus són mutants o l'estat en què cap d'ells ho és, ens podem preguntar: si partim de l'estat en què hi ha  $k$  individus mutants, quina és la probabilitat de què al final tota la població estigui formada per individus d' $A$ , és a dir, formada exclusivament per mutants? I que ho estigui per individus de  $B$ ? Aquestes probabilitats s'anomenen *probabilitat de fixació*, que la denotarem com a  $P_k$ , i d'*extinció*, respectivament. A partir d'ara, quan parlem de probabilitat de fixació utilitzarem  $\rho = P_1$ , és a dir,  $\rho$  ens indica la probabilitat d'arribar a l'estat  $k = n$  partint de l'estat  $k = 1$ . Aquest últim estat l'associem a una primera mutació espontània. Després, se segueix el procés de Moran. Més endavant veurem que el procés de Moran sobre grafs regulars en què  $r > 1$ , té una probabilitat de fixació associada

$$\rho = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^n}}.$$

Aquesta  $\rho$  és la probabilitat de fixació considerada canònica, és a dir, és la que fa de “model” a l'hora de comparar entre les probabilitats de fixació associades a diferents grafs. Suposem que tenim un graf  $G$  amb probabilitat de fixació  $\rho_G$ . Si  $\rho_G > \rho = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^n}}$  direm que  $G$  és un *amplificador* de la selecció ja que afavoreix la fixació dels mutants a la població. En canvi, si  $\rho_G < \rho = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^n}}$  direm que  $G$  és un *supressor* de la selecció ja que afavoreix l'extinció dels mutants.

Per entendre l'evolució dels grafs evolutius ens serà d'utilitat simular el procés de Moran mitjançant cadenes de Márkov. Cada estat de la cadena serà representat per un vèrtex. Hi ha dues maneres de fer-ho:

1. *Cadena de Márkov per nombre de mutants.* (Veure Figura 2.1) Hi ha tants estats a la cadena com individus al graf. En aquesta cadena analitzem el procés estocàstic utilitzant com a paràmetre el nombre total d'agents infectats i l'objectiu és estudiar com s'apropa o s'allunya de 0 i de  $n$ , els estats absorbents. Les probabilitats d'anar d'un estat a un altre són:

- $p_k$  és la probabilitat de què la població de mutants augmenti en un individu; passaria d'haver  $k$  mutants a haver-n'hi  $k + 1$ . És a dir, aniríem de l'estat  $k$  a l' $k + 1$ .
- $q_k$  és la probabilitat de què la població de mutants disminueixi en un individu; passaria d'haver  $k$  mutants a haver-n'hi  $k - 1$ . És a dir, aniríem de l'estat  $k$  a l' $k - 1$ .
- $s_k = 1 - p_k - q_k$ .  $s_k$  és la probabilitat de què el nombre de mutants no variï, és a dir, romanguí havent  $k$ .

El gran avantatge d'aquest model és que per calcular la probabilitat de fixació podem considerar que el nombre de mutants a cada pas sempre augmenta o disminueix en 1, mai roman constant. Malgrat això, té un petit inconvenient: no podem utilitzar-lo sempre. Aquest fet és conseqüència de què sovint no podem donar amb facilitat un valor per a  $p_k$  i altre per a  $q_k$ . Això succeeix quan la probabilitat de passar d'un estat a un altre depèn del conjunt de vèrtexs mutants. Dit amb altres paraules, quan la probabilitat de transició entre estats depèn del conjunt d'individus mutants que hi hagi en aquell moment. De fet, això depèn de la configuració dels vèrtexs del graf. Més endavant veurem que aquest model de cadena de Márkov s'adiu amb grafs  $\Delta$ -regulars. Ara, per calcular la probabilitat de fixació mitjançant aquest model normalitzem les probabilitats acabades d'esmentar de la següent manera:

$$p'_k = \frac{p_k}{p_k + q_k}, \quad q'_k = \frac{q_k}{p_k + q_k},$$

$$s'_k = \frac{s_k}{p_k + q_k} = \frac{1 - (p_k + q_k)}{p_k + q_k} = \frac{1}{p_k + q_k} - 1.$$

Així, tindrem  $p'_k + q'_k = 1$  i podrem considerar  $s'_k = 0$ . Això no ho podrem fer servir quan calculem el temps de fixació ja que en aquest cas és important considerar el temps que el procés passa en un mateix estat.

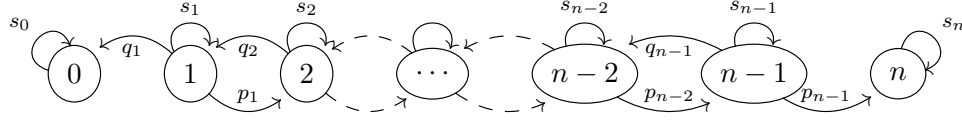


Figura 2.1: Cadena de Màrkov per nombre de mutants amb  $n$  estats.

2. *Cadena de Màrkov per configuracions.* Hi ha tants estats com subconjunts d'individus: cada estat de la cadena representa un subconjunt de mutants. Per tant, el nombre de vèrtexs de la cadena depèn de l'estructura en què estiguin els vèrtexs dins del graf evolutiu. La cadena de Màrkov pot arribar a tenir  $2^n$  vèrtexs i els estats absorbents seran  $\{\emptyset\}$  i  $\{V\}$ . Donats dos subconjunts  $C, D \subseteq 2^n$ , si l'estat  $C$  té una aresta directa a l'estat  $D$ , en un pas podem passar d'un conjunt de mutants  $C$  a un conjunt de mutants  $D$ . Per tant,  $|D| = |C| \pm 1$ . Podem veure un exemple d'això a la Figura 2.4.

## 2.2 Modelització estocàstica del procés de Moran en grafs evolutius

Anem a veure els dos tipus de cadenes descrits mitjançant un exemple. Considerem un graf evolutiu amb tres vèrtexs:  $a$ ,  $b$  i  $c$  connectats tal com mostra la Figura 2.2 on els mutants tenen adaptació genètica  $r > 0$ .

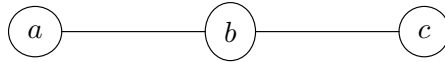


Figura 2.2: Graf evolutiu de tres vèrtexs (recta).

Per estudiar com poden propagar-se les mutacions en la població, podem definir la cadena de Màrkov per nombre d'individus infectats, tal com veiem a la Figura 2.3. L'estat 1, per exemple, inclou tres possibles escenaris: o bé l'individu  $a$  és un mutant, o bé el  $b$  ho és o bé el  $c$ . Modelitzant el procés de Moran pel graf evolutiu de la Figura 2.2 tenim una cadena de Màrkov



amb quatre estats corresponents a 0, 1, 2 i 3 individus infectats. En aquesta cadena hem indicat les probabilitats de passar d'un estat a l'altre,  $p_k$  i  $q_k$ , i de romandre en un mateix,  $s_k$ .

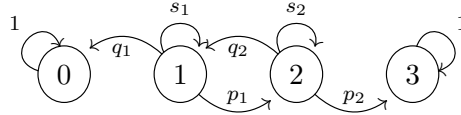


Figura 2.3: Cadena de Màrkov per nombre de mutants associada al graf evolutiu que mostra la Figura 2.2.

Per altra banda, podem definir tots els estats possibles partint de l'estat en què cap individu és mutant. Si aparegués un mutant, aquest podria ser qualsevol dels tres éssers del conjunt i podria infectar als seus veïns. Una vegada tinguéssim dos individus infectats, o bé retrocedim a l'estat en què hi ha un únic mutant o bé l'individu genèticament sa és infectat. Si representem aquest procés indicant quin és el subconjunt de mutants, obtenim la cadena de Màrkov que veiem a la Figura 2.4. En aquesta cadena, cada vèrtex denota el subconjunt de vèrtexs mutants. Podem calcular quines són les probabilitats de passar d'un subconjunt mutant a un altre de la següent manera

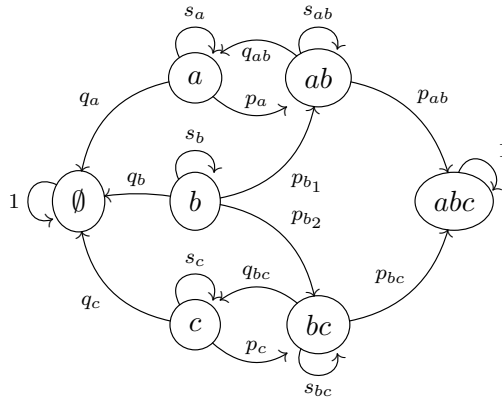


Figura 2.4: Cadena de Màrkov per configuracions associada al graf evolutiu que mostra la Figura 2.2.

$$\begin{aligned}
p_a = p_c &= \frac{r}{r+2} \cdot 1, & q_a = q_c &= \frac{1}{r+2} \cdot \frac{1}{2}, \\
p_b = p_{b_1} + p_{b_2} &= 2 \cdot \left( \frac{r}{r+2} \cdot \frac{1}{2} \right), & q_b &= \frac{2}{r+2} \cdot 1, \\
p_{ab} = p_{bc} &= \frac{r}{2r+1} \cdot \frac{1}{2}, & q_{ab} = q_{bc} &= \frac{1}{2r+1} \cdot 1,
\end{aligned}$$

$$\begin{aligned}
s_a = s_c &= 1 - \left( \frac{r}{r+2} + \frac{1}{2(r+2)} \right), \\
s_b &= 1 - \left( \frac{r}{r+2} + \frac{2}{r+2} \right), \\
s_{ab} = s_{bc} &= 1 - \left( \frac{r}{2(2r+1)} + \frac{1}{2r+1} \right).
\end{aligned}$$

Notem que les probabilitats  $p_C$ ,  $q_C$  i  $s_C$  on  $C \subset \mathcal{P}(V)$  que acabem de calcular són diferents a les probabilitats  $p_k$ ,  $q_k$  i  $s_k$  esmentades anteriorment on  $k \in \{0, \dots, 3\}$ . Malgrat això, podem establir una relació entre elles. En aquest cas particular tenim

$$\begin{aligned}
p_1 &= p_a + p_b + p_c, & p_2 &= p_{ab} + p_{bc}, \\
q_1 &= q_a + q_b + q_c, & q_2 &= q_{ab} + q_{bc}, \\
s_1 &= s_a + s_b + s_c, & s_2 &= s_{ab} + s_{bc}.
\end{aligned}$$

## Capítol 3

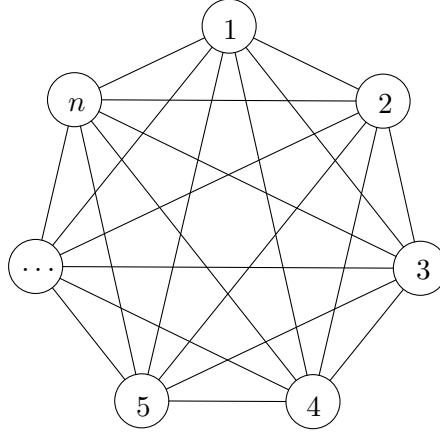
# Probabilitat de fixació

En aquest apartat enunciaré i demostraré algunes propietats sobre la probabilitat de fixació. Començarem amb l'estudi del cas més senzill en què l'adaptació genètica és  $r = 1$  i evolucionarem al cas en què  $r > 0$ . Per començar, estudiarem el procés de mutació d'un graf evolutiu en particular: el clique.

### 3.1 Procés neutral de Moran

Volem calcular quina és la probabilitat de fixació d'una població on tots els individus estan connectats entre si (Figura 3.1) i l'adaptació genètica de cadascun és  $r = 1$ . Estudiarem com es desenvolupa aquest conjunt d'individus seguint el procés de Moran.

Aquest cas és anomenat *procés neutral de Moran* ja que ni els individus mutants ni els no mutants tenen avantatge sobre els altres (parlant en termes reproductius). Sobre aquest graf evolutiu podem raonar utilitzant la cadena de Márkov per nombre de mutants. Això és degut a que estem estudiant el cas en què tenim un graf regular i no ens importa quin és el subconjunt de vèrtexs infectat ja que tots els individus poden convertir-se en mutants de manera equiprobable. Notem que la cadena de Márkov per nombre de mutants associada al graf evolutiu de la Figura 3.1 és la de la Figura 2.1, també amb  $n$  estats.

Figura 3.1: Graf evolutiu de  $n$  vèrtexs (clique).

Per conèixer la probabilitat de què tots els individus es converteixin en mutants partint de l'aparició esporàdica d'un infectat, és a dir, la probabilitat de fixació, calcularem primer la probabilitat de què el nombre de mutants augmenti o disminueixi en un. Començarem assumint que a temps 0 neix espontàniament un mutant a un dels vèrtexs.

- $p_1 = ?$  D'entre els  $n$  individus, hem de triar el mutant per a que es reproduïxi i, dels veïns, qualsevol individu no-mutant per a que sigui substituït per la còpia del primer.

$$p_1 = \frac{1}{n} \cdot \frac{n-1}{n-1} = \frac{1}{n}.$$

- $q_1 = ?$  D'entre els  $n$  individus, hem de triar un no-mutant per a que es reproduïxi i, dels veïns, el mutant per a que sigui substituït per la còpia del primer.

$$q_1 = \frac{n-1}{n} \cdot \frac{1}{n-1} = \frac{1}{n}.$$

- $s_1 = ?$  D'entre els  $n$  individus, hem de triar un individu per a que es reproduïxi i, dels veïns, un altre de la mateixa espècie (mutant - mutant o bé no mutant - no mutant) per a que sigui substituït per la còpia del primer. Com només hi ha un mutant, la única combinació possible és

triar dos no mutants.

$$s_1 = \frac{n-1}{n} \cdot \frac{n-2}{n-1} = \frac{n-2}{n} = 1 - \frac{2}{n}.$$

També, com generalment farem, podem calcular  $s_1$  sabent que

$$s_1 = 1 - (p_1 + q_1) = 1 - \frac{2}{n}.$$

Anàlogament obtenim aquestes probabilitats pel cas genèric en què hi ha  $k$  mutants dins la població:

$$\begin{aligned} p_k &= \frac{k}{n} \cdot \frac{(n-1) - (k-1)}{n-1} = \frac{k}{n} \cdot \frac{n-k}{n-1}, \\ q_k &= \frac{n-k}{n} \cdot \frac{k}{n-1}, \\ s_k &= 1 - p_k - q_k. \end{aligned}$$

Notem que  $p_k = q_k$  ja que l'adaptació genètica dels individus és 1: donats  $k$  mutants, que un no-mutant muti i que un mutant es curi són successos equiprobables.

Per simplificar aquest exemple imposarem  $s_k = 0$ . I, per tant,  $p_k = q_k = \frac{1}{2}$ . Definim com a  $P_k$  la probabilitat de fixació quan hi ha  $k$  mutants. Conseqüentment, tenim que

$$P_k = \frac{1}{2}P_{k-1} + \frac{1}{2}P_{k+1}.$$

Sabem que  $P_n = 1$  i  $P_0 = 0$ . Aleshores:

$$\begin{aligned} P_{n-1} &= \frac{1}{2}P_{n-2} + \frac{1}{2}, \\ P_1 &= \frac{1}{2}P_2, \\ P_2 &= \frac{1}{2}P_1 + \frac{1}{2}P_3 = \frac{1}{2} \cdot \left(\frac{1}{2}P_2\right) + \frac{1}{2}P_3 = \frac{1}{4}P_2 + \frac{1}{2}P_3 \Rightarrow P = \frac{4}{3} \cdot \frac{1}{2}P_3 = \frac{2}{3}P_3, \\ P_3 &= \frac{1}{2}P_2 + \frac{1}{2}P_4 = \frac{1}{2} \cdot \left(\frac{2}{3}P_3\right) + \frac{1}{2}P_4 = \frac{1}{3}P_3 + \frac{1}{2}P_4 \Rightarrow P_3 = \frac{3}{2} \cdot \frac{1}{2}P_4 = \frac{3}{4}P_4. \end{aligned}$$

Veurem per inducció que

$$P_k = \frac{k}{k+1} P_{k+1}.$$

Suposem que és cert fins a  $n-2$ . Vegem què passa per  $n-1$ :

$$\begin{aligned} P_{n-1} &= \frac{1}{2} P_{n-2} + \frac{1}{2} = \frac{1}{2} \cdot \frac{n-1}{n-2} P_{n-1} + \frac{1}{2}. \\ \Rightarrow \left(1 - \frac{1}{2} \cdot \frac{n-2}{n-1}\right) P_{n-1} &= \frac{2(n-1) - (n-2)}{2(n-1)} P_{n-1} \\ &= \frac{n}{2(n-1)} P_{n-1} = \frac{1}{2}. \\ \Rightarrow P_{n-1} &= \frac{1}{2} \cdot \frac{2(n-1)}{n} = \frac{n-1}{n} = \frac{n-1}{n} P_n. \end{aligned}$$

Ara, ja hem vist que  $P_k = \frac{k}{k+1} P_{k+1}$ . A més, per  $k = n-1$  tenim que  $P_{n-1} = \frac{n-1}{n}$  i si substituïm aquest resultat a la primera equació, obtenim el següent:

$$\begin{aligned} P_{n-1} &= \frac{n-1}{n}, \\ P_{n-2} &= \frac{n-2}{n-1} P_{n-1} = \frac{n-2}{n-1} \cdot \frac{n-1}{n} = \frac{n-2}{n}, \\ P_{n-3} &= \frac{n-3}{n-2} P_{n-2} = \frac{n-3}{n-2} \cdot \frac{n-2}{n} = \frac{n-3}{n}, \\ &\vdots \\ P_k &= \frac{k}{k+1} P_{k+1} = \frac{k}{k+1} \cdot \frac{k+1}{n} = \frac{k}{n}. \end{aligned}$$

Per tant, concloem que la probabilitat de fixació per a qualsevol  $k$  és

$$P_k = \frac{k}{n}.$$

I, doncs, la probabilitat de fixació per a  $k = 1$  és

$$\rho = \frac{1}{n}.$$

A la següent secció veurem que per a qualsevol graf evolutiu seguint el

procés neutral de Moran,  $r = 1$ , la probabilitat de fixació és sempre la mateixa,  $\rho = \frac{1}{n}$ .

### 3.2 Propietats de la probabilitat de fixació

Amb els següents resultats, extrets de [2], veurem que podem donar fites per la probabilitat de fixació. El primer lema que enunciaré i demostrareu ens diu que, per  $r = 1$ , la probabilitat de fixació és  $\frac{1}{n}$  independentment de la topologia del graf.

**Lema 2.** Sigui  $G = (V, E)$  un graf connex no dirigit amb  $n$  vèrtexs. Per tant, pel procés neutral de Moran, és a dir, per  $r = 1$ , tenim que  $\rho = \frac{1}{n}$ .

*Demostració.* Considerem una variant del procés on cada vèrtex comença amb un color (dels dos possibles: un representarà als mutants i l'altre als no-mutants) i cada vèrtex té  $r = 1$ . Ara, seguirem el procés de Moran: a cada pas un vèrtex és escollit uniformement de manera aleatòria i el seu color es propaga a algun dels seus veïns escollit, també, aleatòriament. Per tant, algun dels dos colors acabarà expandint-se sobre tot el graf amb probabilitat 1. I la probabilitat de què el color inicial del vèrtex  $x$  sigui el color que acabarà dominant el graf és

$$\rho = P_1 = \frac{1}{n} \sum_{x \in V} P_1(x) = \frac{1}{n}$$

on  $P_1(x)$  és la probabilitat de què el vèrtex  $x$  propagui el seu color a tot el graf i, al ser  $P_1(x)$  una distribució de probabilitats,  $\sum_{x \in V} P_1(x) = 1$ .  $\square$

A continuació, veurem una relació entre les probabilitats de fixació de dos conjunts pel procés neutral de Moran.

**Lema 3.** Donat un graf  $G = (V, E)$  no dirigit i connex i siguin  $C, D \subseteq V$  dos conjunts on  $r = 1$ . Aleshores, tenim que  $P_{C \cup D} = P_C + P_D - P_{C \cap D}$ .

*Demostració.* Com hem fet abans, considerarem que cada vèrtex té el seu propi color i que  $r = 1$ . En qualsevol moment podem considerar que un subconjunt de colors, que denotarem com a  $C$ , representa la població de mutants. És a dir, tots els vèrtexs pintats amb un color dels que conté  $C$  seran considerats

mutants. Per tant, tots aquests vèrtexs pertanyeran a  $C$ . Ara, doncs, per qualsevol  $C \subseteq V$ ,  $P_C$  és la probabilitat de què només romanguin els colors de  $C$ . Aleshores, tenim

$$P_{C \cup D} = \sum_{x \in C \cup D} P_1(x) = \sum_{x \in C} P_1(x) + \sum_{x \in D} P_1(x) - \sum_{x \in C \cap D} P_1(x) = P_C + P_D - P_{C \cap D}$$

on  $P_1(x)$  és la probabilitat de què el vèrtex  $x$  propagui el seu color a tot el graf.  $\square$

Els següent enunciat fita superiorment la probabilitat de fixació per a grafs evolutius amb  $r > 0$ .

**Lema 4.** Sigui  $G = (V, E)$  un graf no dirigit de  $n$  vèrtexs. Aleshores

$$\rho \leq 1 - \frac{1}{n+r}$$

per qualsevol  $r > 0$ .

*Demostració.* Per donar una fita superior de  $\rho$  el que farem és assumir que la fixació s'assoleix quan hi ha dos mutants, és a dir, es donarà abans que en una cadena de Márkov usual amb  $n$  estats.

$\forall x \in V$  definim  $Q(x) = \sum_{xy \in E} \frac{1}{\deg(y)}$ , per tant  $\sum_{x \in V} Q(x) = n$ . Ara, suposem que hi ha un únic mutant. Es donen els següents resultats:

- La probabilitat de què un nou mutant es creï és  $p_1(x) = \frac{r}{n-1+r}$ .
- La probabilitat de què algun dels veïns del mutant sani al mutant és  $q_1(x) = \frac{1}{n-1+r} Q(x)$ .
- La probabilitat de què el procés ramangui en el mateix estat ve donada per  $s_1(x) = 1 - p_1(x) - q_1(x)$ .

La probabilitat de què la població de mutants estigui formada per dos individus (és a dir, que el primer canvi d'estat sigui la creació d'un nou mutant) ve donada per

$$P(x) = \frac{p_1(x)}{p_1(x) + q_1(x)} = \frac{r}{r + Q(x)}.$$



En conseqüència, la probabilitat de fixació segons el nou procés que hem definit és

$$\tilde{\rho} = \frac{1}{n} \sum_{x \in V} P(x) = \frac{r}{n} \sum_{x \in V} \frac{1}{r + Q(x)}.$$

Reescrivint  $\tilde{\rho} = \frac{r}{n} \sum_{k=1}^n (r + Q_k)^{-1}$  volem trobar el màxim valor de  $\tilde{\rho}$  sabent que  $Q_k > 0 \forall k$  i que  $\sum_{k=1}^n Q_k = \sum_{x \in V} Q(x) = n$ . Si fem més dèbil la primera restricció i suposem  $Q_k \geq 0$ , la suma és màxima quan  $Q_1 = n$  i  $Q_2 = \dots = Q_n = 0$ . Aleshores,

$$\begin{aligned} \rho \leq \tilde{\rho} &\leq \frac{r}{n} \left( \frac{1}{r+n} + (n-1) \frac{1}{r+0} \right) = \frac{r}{n} \left( \frac{1}{r+n} + \frac{n}{r} - \frac{1}{r} \right) \\ &= 1 + \frac{r}{n} \left( \frac{1}{r+n} - \frac{1}{r} \right) = 1 + \frac{r}{n(r+n)} - \frac{1}{n} = 1 + \frac{r - (r+n)}{n(r+n)} \\ &= 1 - \frac{1}{r+n}. \end{aligned}$$

□

## Capítol 4

# Procés de Moran per a grafs connexos no dirigits i $\Delta$ -regulars

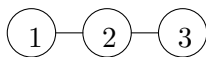
Donat un graf  $G$  connex no dirigit, es defineix com a  $W$  la seva *matriu de mutació* on  $W = [w_{ij}]$  i  $w_{ij}$  és la probabilitat de què l'individu  $i$  substitueixi a l'individu  $j$ .

$$w_{ij} = \begin{cases} \frac{1}{\deg(i)}, & \text{si } i \text{ i } j \text{ estan connectats,} \\ 0, & \text{altrament.} \end{cases}$$

Notem que cada aresta té el mateix pes.

Veiem alguns exemples associats a diferents grafs:

- Recta. Escollim un graf amb forma de recta formada per tres punts.

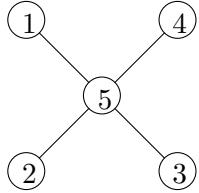


La matriu de mutació corresponent és

$$W = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}.$$

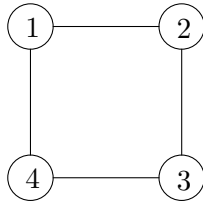
- Estrella. Escollim una estrella formada per 4+1 vèrtexs.

La matriu de mutació associada és



$$W = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \end{pmatrix}.$$

- Quadrat. Escollim un graf amb forma de quadrat tal com mostra la següent figura.

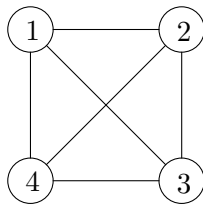


La matriu de mutació corresponent és

$$W = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Notem que la matriu associada a aquest graf evolutiu és doblement estocàstica. Això és perquè cada vèrtex està connectat amb altres dos vèrtexs, és a dir, cada individu pot ser influenciat per altres dos de manera equiprobable. Aquest fet es dona per a tot graf  $\Delta$ -regular, tal com veurem al lema 5.

- $\mathbb{K}_4$ . Escollim un graf evolutiu 3-regular amb quatre vèrtexs.



La matriu de mutació associada és

$$W = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}.$$

Observem que la matriu de mutació associada al graf  $\mathbb{K}_4$  també és doblement estocàstica. La interpretació és anàloga al cas del quadrat.

A continuació veurem el que hem pogut observar en els dos últims exemples: la matriu de mutació associada a un graf regular és doblement estocàstica.

**Lema 5.** Donat un graf  $G = (V, E)$  amb  $|V| = n$ ,  $G$  és regular  $\implies W_G$  és doblement estocàstica.

*Demostració.*  $G$  és regular  $\implies W_G$  és doblement estocàstica. Suposem que  $G$  és  $\Delta$ -regular. Aleshores,

$$\forall i = 1, \dots, n \quad \sum_{j=1}^n w_{ij} = \Delta \cdot \frac{1}{\Delta} = 1.$$

$$\forall j = 1, \dots, n \quad \sum_{i=1}^n w_{ij} = \Delta \cdot \frac{1}{\Delta} = 1.$$

□

Ara, podem utilitzar els conceptes que hem definit per expressar la probabilitat de fixació en funció del conjunt d'individus infectats. Anomenarem  $P_C$  a la probabilitat de fixació donat un conjunt  $C$  de mutants. Com ja hem vist en diversos exemples, podem expressar  $P_C$  en funció de la probabilitat de fixació de dos estats:

$$\begin{cases} |C| \rightarrow |C| + 1, \text{ la població de mutants augmenta: } C \rightarrow C \cup \{j\} \text{ on } j \notin C, \\ |C| \rightarrow |C| - 1, \text{ la població de mutants disminueix: } C \rightarrow C \setminus \{i\} \text{ on } i \in C. \end{cases}$$

Així doncs, expressarem  $P_C$  en funció de  $P_{C \cup \{j\}}$  i de  $P_{C \setminus \{i\}}$  tenint en compte quina és la probabilitat d'arribar als dos estats partint de  $C$ . Per facilitar els càlculs, definirem  $W_C$  com l'adaptació genètica total de  $C$ .

$$\begin{aligned} P(C \rightarrow C \cup \{j\}) &= \sum_{i \text{ mutant}} \sum_{\substack{j \text{ no} \\ \text{mutant}}} \frac{r \cdot w_{ij}}{W_C} = \sum_{i \in C} \sum_{j \notin C} \frac{r \cdot w_{ij}}{W_C}, \\ P(C \rightarrow C \setminus \{i\}) &= \sum_{i \text{ mutant}} \sum_{\substack{j \text{ no} \\ \text{mutant}}} \frac{w_{ji}}{W_C} = \sum_{i \in C} \sum_{j \notin C} \frac{w_{ji}}{W_C}. \end{aligned}$$

Ara, notem que

$$W_C = r \cdot \sum_{i \in C} \sum_{j \notin C} w_{ij} + \sum_{j \notin C} \sum_{i \in C} w_{ji} = \sum_{i \in C} \sum_{j \notin C} (r w_{ij} + w_{ji})$$

on  $(i, j) \in E$  són les arestes que surten de  $C$  i les arestes  $(j, i) \in E$  són les que entren a  $C$ . Podem veure un exemple a la Figura 4.1. En aquest exemple, com existeix l'aresta  $(i, j) \in E$  amb  $i \in C$  i  $j \notin C$ ,  $i$  pot infectar a  $j$  amb probabilitat  $w_{ij}$  i  $j$  pot curar a  $i$  amb probabilitat  $w_{ji}$ .

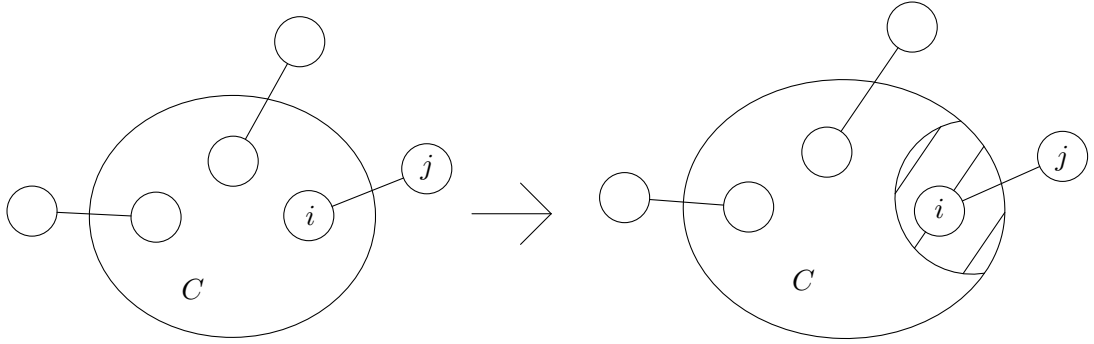


Figura 4.1: Possible evolució d'un graf evolutiu en un pas:  $j$  cura a  $i$ .

D'aquesta manera obtenim la següent expressió:

$$P_C = \frac{\sum_{i \in C} \sum_{j \notin C} (r w_{ij} P_{C \cup \{j\}} + w_{ji} P_{C \setminus \{i\}})}{\sum_{i \in C} \sum_{j \notin C} (r w_{ij} + w_{ji})}. \quad (4.1)$$

El denominador pren el valor de l'adaptació genètica que té la població en aquell estat. I el numerador és la suma de probabilitats de tots els estats als que podem arribar partint de  $C$ .

Altre concepte que ens serà d'utilitat és el de temperatura. La *temperatura* del vèrtex  $i$  és

$$T_i = \sum_{j \in V} w_{ji}.$$

La temperatura d'un vèrtex ens indica amb quina freqüència aquest vèrtex és reemplaçat, és a dir, un *node calent* és reemplaçat molt sovint i un *node fred* rarament és substituït.

En el cas de grafs  $\Delta$ -regulars tenim  $T_i = \sum_{j \in V} w_{ji} = \frac{1}{\Delta}$ .

Direm que un graf és isoterm quan tots els vèrtexs tenen la mateixa temperatura. Ara, introduïrem un teorema que relaciona l'estructura d'un graf amb la temperatura dels seus nodes (teorema 3.1 de [1]).

**Teorema 6.** *Donat un graf  $G=(V, E)$  connex no dirigit i sigui  $r$  qualsevol, tenim que  $G$  és regular  $\iff G$  és isoterm.*

*Demostració.*  $G$  és regular  $\Rightarrow G$  és isoterm. Suposem que  $G$  és  $\Delta$ -regular. Aleshores tenim que

$$T_i = \sum_{j \in V} w_{ji} = \frac{1}{\Delta} = \sum_{i \in V} w_{ji} = T_j.$$

Es a dir,  $\forall i, j$  tenim  $T_i = T_j$ . Conseqüentment,  $G$  és isoterm.

$G$  és isoterm  $\Rightarrow G$  és regular. Suposem que no és cert per arribar a contradicció. Considerem el conjunt  $S = \{i \mid \deg(i) = \min\{\deg(v) \mid v \in V\}\}$ . Per hipòtesi  $S \neq V$ . Aleshores, existeix un vèrtex  $i \in S$  que està connectat al vèrtex  $j \in V \setminus S$ . En conseqüència,

$$\sum_v w_{vi} = w_{ji} + \sum_{v \neq j} w_{vi} < \frac{1}{\deg(i)} + \sum_{v \neq j} w_{vi} \leq \frac{1}{\deg(i)} + \frac{\deg(i) - 1}{\deg(i)} = 1$$

arribem a contradicció, tal com volíem veure.  $\square$

El següent enunciat és el que ens permetrà calcular les probabilitats de fixació de grafs regulars amb  $r > 1$  sense la necessitat de fer tots els càlculs que hem realitzat anteriorment.

**Corol·lari 7.** Sigui  $G=(V, E)$  un graf isoterm no dirigit amb  $|V| = n$  on l'adaptació genètica dels mutants és  $r > 1$ . Aleshores, la probabilitat de fixació associada és

$$\rho = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^n}}.$$

Aquesta és l'anomenada probabilitat de fixació canònica ja que és la que respon a la majoria dels grafs isoterms.

*Demostració.* Mitjançant el teorema anterior, com  $G$  és un graf isoterm sabem que també és regular. Suposarem que és  $\Delta$ -regular. Per altra banda, utilitzarem la fórmula (4.1) però denotarem  $P_k = P_C$  sent  $|C| = k$ . Aleshores

$$P_k = \frac{a \cdot \left(\frac{1}{\Delta} \cdot r \cdot P_{k+1} + \frac{1}{\Delta} \cdot P_{k-1}\right)}{a \cdot \left(\frac{1}{\Delta} \cdot r + \frac{1}{\Delta}\right)} = \frac{r}{r+1} \cdot P_{k+1} + \frac{1}{r+1} \cdot P_{k-1}$$

on  $a$  és el nombre d'arestes  $(i, j) \in E$  amb  $i \in C$  i  $j \notin C$ .

Al ser una equació en diferències buscarem una solució de la forma  $P_k = \theta^k$  (tal com podem veure a la secció 3.9 de [4]).

$$\begin{aligned} \theta^k &= \frac{r}{r+1} \cdot \theta^{k+1} + \frac{1}{r+1} \cdot \theta^{k-1}, \\ \xrightarrow{(\theta^{k-1})^{-1}} \frac{r}{r+1} \cdot \theta^2 - \theta + \frac{1}{r+1} &= 0. \end{aligned}$$

Aquesta equació de segon grau té com a solucions

$$\theta_1 = 1 \text{ i } \theta_2 = \frac{1}{r}.$$

En conseqüència la solució general de l'equació de  $P_k$  és  $P_k = A \theta_1^k + B \theta_2^k$  per constants arbitràries  $A$  i  $B$ . Utilitzem les condicions de vora,  $P_0 = 0$  i  $P_n = 1$ , per trobar-les:

$$\begin{aligned} 0 &= P_0 = A + B \Rightarrow A = -B, \\ 1 &= P_n = A \theta_1^n + B \theta_2^n = A \cdot 1 + B \cdot \frac{1}{r^n} = B \cdot \left(\frac{1}{r^n} - 1\right). \\ \Rightarrow B &= \left(\frac{1}{r^n} - 1\right)^{-1} = \left(\frac{1 - r^n}{r^n}\right)^{-1} = \frac{r^n}{1 - r^n}. \\ \Rightarrow A &= \frac{-r^n}{1 - r^n}. \end{aligned}$$

$$\begin{aligned}
 P_k &= \frac{-r^n}{1-r^n} \cdot 1^k + \frac{r^n}{1-r^n} \cdot \left(\frac{1}{r}\right)^k = \frac{-r^n \cdot r^k + r^n}{(1-r^n) \cdot r^k} \\
 &= \frac{r^n \cdot (1-r^k)}{r^k \cdot (1-r^n)} = \frac{\frac{(1-r^k)}{r^k}}{\frac{(1-r^n)}{r^n}} \\
 &= \frac{1 - \frac{1}{r^k}}{1 - \frac{1}{r^n}}.
 \end{aligned}$$

I, per tant

$$\rho = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^n}},$$

tal com volíem veure.

□



## Capítol 5

### L'estrella

Volem calcular quina és la probabilitat de fixació d'una població de  $n + 1$  individus connectats entre si tal com mostra la Figura 5 i on l'adaptació genètica de cadascun és  $r$ .

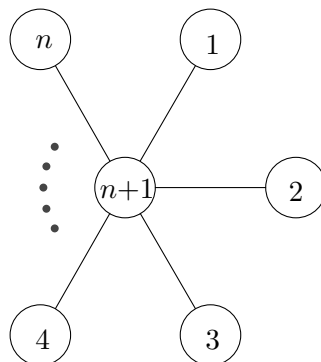


Figura 5.1: Graf evolutiu amb format d'estrella de  $n + 1$  vèrtexs.

Notem que cal distingir el cas en què el centre està ocupat per un mutant del cas en què està ocupat per un no-mutant. Això ve de què un individu que està a les fulles de l'estrella només està influenciat per l'individu del centre mentre que l'individu del centre està influenciat per la resta d'individus. D'aquesta manera definirem les probabilitats de fixació següent:

- $P_k^{\text{on}}$  és la probabilitat de fixació havent  $k$  mutants a les fulles i un mutant al centre.

- $P_k^{\text{off}}$  és la probabilitat de fixació havent  $k$  mutants a les fulles i cap mutant al centre.

Abans de calcular aquestes probabilitats, definirem l'adaptació genètica total del graf per a  $k$  mutants com

$$\begin{aligned} W_k^{\text{on}} &= (n - k) + (k + 1) \cdot r \\ W_k^{\text{off}} &= (n + 1 - k) + k \cdot r \end{aligned}$$

per simplificar els pròxims càlculs. També, anàlogament al cas general, utilitzarem  $p_k^{\text{on}}$ ,  $p_k^{\text{off}}$ ,  $q_k^{\text{on}}$  i  $q_k^{\text{off}}$ . Aleshores, obtenim

$$P_k^{\text{on}} = q_k^{\text{on}} \cdot P_k^{\text{off}} + p_k^{\text{on}} \cdot P_{k+1}^{\text{on}} + (1 - (q_k^{\text{on}} + p_k^{\text{on}})) \cdot P_k^{\text{on}} \quad (5.1)$$

$$P_k^{\text{off}} = p_k^{\text{off}} \cdot P_k^{\text{on}} + q_k^{\text{off}} \cdot P_{k-1}^{\text{off}} + (1 - (q_k^{\text{off}} + p_k^{\text{off}})) \cdot P_k^{\text{off}} \quad (5.2)$$

on

$$\begin{aligned} q_k^{\text{on}} &= \frac{n - k}{W_k^{\text{on}}} \cdot 1 = \frac{n - k}{W_k^{\text{on}}}, & q_k^{\text{off}} &= \frac{1}{W_k^{\text{off}}} \cdot \frac{k}{n} = \frac{k}{n \cdot W_k^{\text{off}}}, \\ p_k^{\text{on}} &= \frac{r}{W_k^{\text{on}}} \cdot \frac{n - k}{n} = \frac{(n - k) \cdot r}{n \cdot W_k^{\text{on}}}, & p_k^{\text{off}} &= \frac{k \cdot r}{W_k^{\text{off}}} \cdot 1 = \frac{k \cdot r}{W_k^{\text{off}}}. \end{aligned}$$

Ara, normalitzarem les equacions (5.1) i (5.2) per obtenir una expressió de  $P_k^{\text{on}}$  i de  $P_k^{\text{off}}$  que obviï els bucles.

(5.1) Dividim la primera equació per  $q_k^{\text{on}} + p_k^{\text{on}}$ :

$$\begin{aligned} q_k^{\text{on}} + p_k^{\text{on}} &= \frac{n - k}{W_k^{\text{on}}} + \frac{(n - k) \cdot r}{n \cdot W_k^{\text{on}}} = \frac{n \cdot (n - k) + r \cdot (n - k)}{n \cdot W_k^{\text{on}}}, \\ \frac{q_k^{\text{on}}}{q_k^{\text{on}} + p_k^{\text{on}}} &= \frac{n \cdot (n - k)}{n \cdot (n - k) + r \cdot (n - k)} = \frac{n}{n + r}, \\ \frac{p_k^{\text{on}}}{q_k^{\text{on}} + p_k^{\text{on}}} &= \frac{r \cdot (n - k)}{n \cdot (n - k) + r \cdot (n - k)} = \frac{r}{n + r}. \end{aligned}$$

$$\begin{aligned}
\frac{1}{q_k^{\text{on}} + p_k^{\text{on}}} \cdot P_k^{\text{on}} &= \frac{q_k^{\text{on}}}{q_k^{\text{on}} + p_k^{\text{on}}} \cdot P_k^{\text{off}} + \frac{p_k^{\text{on}}}{q_k^{\text{on}} + p_k^{\text{on}}} \cdot P_{k+1}^{\text{on}} + \frac{1 - (q_k^{\text{on}} + p_k^{\text{on}})}{q_k^{\text{on}} + p_k^{\text{on}}} \cdot P_k^{\text{on}}, \\
\frac{1}{q_k^{\text{on}} + p_k^{\text{on}}} \cdot P_k^{\text{on}} &= \frac{n}{n+r} \cdot P_k^{\text{off}} + \frac{r}{n+r} \cdot P_{k+1}^{\text{on}} + \frac{1}{q_k^{\text{on}} + p_k^{\text{on}}} \cdot P_k^{\text{on}} + P_k^{\text{on}}, \\
P_k^{\text{on}} &= \frac{n}{n+r} \cdot P_k^{\text{off}} + \frac{r}{n+r} \cdot P_{k+1}^{\text{on}}.
\end{aligned} \tag{5.3}$$

(5.2) Dividim la segona equació per  $q_k^{\text{off}} + p_k^{\text{off}}$ :

$$\begin{aligned}
q_k^{\text{off}} + p_k^{\text{off}} &= \frac{k}{n \cdot W_k^{\text{off}}} + \frac{r \cdot k}{W_k^{\text{off}}} = \frac{k + n \cdot r \cdot k}{n \cdot W_k^{\text{off}}}, \\
\frac{q_k^{\text{off}}}{q_k^{\text{off}} + p_k^{\text{off}}} &= \frac{n \cdot k}{n \cdot (n \cdot r \cdot k + k)} = \frac{1}{n \cdot r + 1}, \\
\frac{p_k^{\text{off}}}{q_k^{\text{off}} + p_k^{\text{off}}} &= \frac{n \cdot r \cdot k}{n \cdot r \cdot k + k} = \frac{n \cdot r}{n \cdot r + 1}.
\end{aligned}$$

$$\begin{aligned}
\frac{1}{q_k^{\text{off}} + p_k^{\text{off}}} \cdot P_k^{\text{off}} &= \frac{p_k^{\text{off}}}{q_k^{\text{off}} + p_k^{\text{off}}} \cdot P_k^{\text{on}} + \frac{q_k^{\text{off}}}{q_k^{\text{off}} + p_k^{\text{off}}} \cdot P_{k-1}^{\text{off}} + \frac{1 - (q_k^{\text{off}} + p_k^{\text{off}})}{q_k^{\text{off}} + p_k^{\text{off}}} \cdot P_k^{\text{off}}, \\
\frac{1}{q_k^{\text{off}} + p_k^{\text{off}}} \cdot P_k^{\text{off}} &= \frac{n \cdot r}{n \cdot r + 1} \cdot P_k^{\text{on}} + \frac{1}{n \cdot r + 1} \cdot P_{k-1}^{\text{off}} + \frac{1 - (q_k^{\text{off}} + p_k^{\text{off}})}{q_k^{\text{off}} + p_k^{\text{off}}} \cdot P_k^{\text{off}}, \\
P_k^{\text{off}} &= \frac{n \cdot r}{n \cdot r + 1} \cdot P_k^{\text{on}} + \frac{1}{n \cdot r + 1} \cdot P_{k-1}^{\text{off}}.
\end{aligned} \tag{5.4}$$

Ara, utilitzant l'equació (5.3) obtenim  $\left\{ \begin{array}{l} P_k^{\text{on}} = \frac{n}{n+r} \cdot P_k^{\text{off}} + \frac{r}{n+r} \cdot P_{k+1}^{\text{on}}, \\ P_{k-1}^{\text{on}} = \frac{n}{n+r} \cdot P_{k-1}^{\text{off}} + \frac{r}{n+r} \cdot P_i^{\text{on}}. \end{array} \right.$

$$\begin{aligned}
\frac{r}{n+r} \cdot P_k^{\text{on}} &= P_{k-1}^{\text{on}} - \frac{n}{n+r} \cdot P_{k-1}^{\text{off}}, \\
\Rightarrow P_k^{\text{on}} &= \frac{n+r}{r} \cdot \left( P_{k-1}^{\text{on}} - \frac{n}{n+r} \cdot P_{k-1}^{\text{off}} \right) = \frac{n}{r} \cdot P_{k-1}^{\text{on}} + P_{k-1}^{\text{on}} - \frac{n}{r} \cdot P_{k-1}^{\text{off}}, \\
\Rightarrow P_k^{\text{on}} &= P_{k-1}^{\text{on}} + \frac{n}{r} \cdot \left( P_{k-1}^{\text{on}} - P_{k-1}^{\text{off}} \right).
\end{aligned} \tag{5.5}$$

Volem expressar  $P_1^{\text{off}}, P_2^{\text{on}}, P_2^{\text{off}}, \dots$  com a funció de  $P_1^{\text{on}}$ . Per fer-ho, utilitzarem (5.4) i (5.5):

$$\begin{aligned}
P_1^{\text{off}} &= \frac{nr}{nr+1} P_1^{\text{on}} + \frac{1}{nr+1} P_0^{\text{off}} = \frac{nr}{nr+1} P_1^{\text{on}}, \\
P_2^{\text{on}} &= P_1^{\text{on}} + \frac{n}{r} (P_1^{\text{on}} - P_1^{\text{off}}) = P_1^{\text{on}} \left( 1 + \frac{n}{r} \left( 1 - \frac{nr}{nr+1} \right) \right) \\
&= P_1^{\text{on}} \left( 1 + \frac{n}{r} \cdot \frac{nr+1-nr}{nr+1} \right) = P_1^{\text{on}} \left( 1 + \frac{n}{r(nr+1)} \right), \\
P_2^{\text{off}} &= \frac{nr}{nr+1} P_2^{\text{on}} + \frac{1}{nr+1} P_1^{\text{off}} \\
&= \frac{nr}{nr+1} \left( 1 + \frac{n}{r(nr+1)} \right) P_1^{\text{on}} + \frac{1}{nr+1} \cdot \frac{nr}{nr+1} P_1^{\text{on}} \\
&= P_1^{\text{on}} \left( \frac{nr}{nr+1} + \frac{n^2 r}{r(nr+1)^2} + \frac{nr}{(nr+1)^2} \right) \\
&= P_1^{\text{on}} \left( \frac{n^2 r^2 + nr + n^2 + nr}{(nr+1)^2} \right) = P_1^{\text{on}} \left( \frac{n^2 r^2 + 2nr + n^2}{(nr+1)^2} \right), \\
P_3^{\text{on}} &= P_2^{\text{on}} + \frac{n}{r} (P_2^{\text{on}} - P_2^{\text{off}}) \\
&= P_1^{\text{on}} \left( 1 + \frac{n}{r(nr+1)} \right) + \frac{n}{r} \left[ P_1^{\text{on}} \left( 1 + \frac{n}{r(nr+1)} \right) - P_1^{\text{on}} \left( \frac{n^2 r^2 + 2nr + n^2}{(nr+1)^2} \right) \right] \\
&= P_1^{\text{on}} \left( 1 + \frac{n}{r(nr+1)} + \frac{n}{r} + \frac{n^2}{r^2(nr+1)} - \frac{n^3 r^2 + 2n^2 r + n^3}{r(nr+1)^2} \right) \\
&= P_1^{\text{on}} \left( 1 + \frac{n^2 r^2 + nr + n^3 r^3 + 2n^2 r^2 + nr + n^3 r + n^2 - n^3 r^3 - 2n^2 r^2 + n^3 r}{r^2(nr+1)^2} \right) \\
&= P_1^{\text{on}} \left( 1 + \frac{n^2 r^2 + nr + n^2 + nr}{r^2(nr+1)^2} \right) = P_1^{\text{on}} \left( 1 + \frac{n}{r(nr+1)} + \frac{n(n+r)}{r^2(nr+1)^2} \right) \\
&= P_1^{\text{on}} \left( 1 + \frac{n}{n+r} \sum_{j=1}^2 \left( \frac{n+r}{r(nr+1)} \right)^j \right).
\end{aligned}$$

Utilitzant inducció podem veure

$$P_k^{\text{on}} = P_1^{\text{on}} \cdot \left( 1 + \frac{n}{n+r} \sum_{j=1}^{k-1} \left( \frac{n+r}{r(nr+1)} \right)^j \right).$$

Ara doncs, aplicant l'expressió que acabem de trobar obtenim:

$$P_{n-1}^{\text{on}} = P_1^{\text{on}} \cdot \left( 1 + \frac{n}{n+r} \sum_{j=1}^{n-2} \left( \frac{n+r}{r(nr+1)} \right)^j \right),$$

$$P_{n-1}^{\text{off}} = \frac{n \cdot r}{n \cdot r + 1} \cdot P_{n-1}^{\text{on}} + \frac{1}{n \cdot r + 1} \cdot P_{n-2}^{\text{off}},$$

$$P_n^{\text{on}} = P_{n-1}^{\text{on}} + \frac{n}{r} \cdot (P_{n-1}^{\text{on}} - P_{n-1}^{\text{off}}).$$

Com  $P_n^{\text{on}} = 1$ , tenim

$$P_1^{\text{on}} = \frac{1}{1 + \frac{n}{n+r} \sum_{j=1}^{n-1} \left( \frac{n+r}{r(nr+1)} \right)^j}.$$

I, per (5.3) i (5.4) tenim  $P_0^{\text{on}} = \frac{r}{r+n}$  i  $P_1^{\text{off}} = \frac{nr}{nr+1} P_1^{\text{on}}$ , respectivament. Aleshores, obtenim que la probabilitat mitjana de fixació és

$$\rho = \frac{n \frac{nr}{nr+1} + \frac{r}{r+n}}{(n+1) \cdot \left( 1 + \frac{n}{n+r} \sum_{j=1}^{n-1} \left( \frac{n+r}{r(nr+1)} \right)^j \right)}.$$

I, si fem el límit quan  $n \rightarrow \infty$  obtenim

$$\begin{aligned} \rho &\approx \lim_{n \rightarrow \infty} \frac{n}{(n+1) \cdot \left( 1 + \sum_{j=1}^{n-1} \left( \frac{n+r}{r(nr+1)} \right)^j \right)} = \lim_{n \rightarrow \infty} \frac{1}{1 + \sum_{j=1}^{n-1} \left( \frac{n+r}{r(nr+1)} \right)^j} \\ &= \lim_{n \rightarrow \infty} \frac{1}{1 + \sum_{j=1}^{n-1} \left( \frac{1}{r^2} \right)^j} = \frac{1}{1 + \frac{r^2 - r^{2n}}{1 - r^2}} = \frac{1 - r^2}{1 - r^2 + r^2 - r^{2n}} = \frac{1 - r^2}{1 - r^{2n}} \\ &= \frac{1 - \frac{1}{r^2}}{1 - \frac{1}{r^{2n}}}. \end{aligned}$$

És a dir que la probabilitat de fixació de l'estrella és semblant a la probabilitat de fixació canònica.

Es diu que l'estrella és un graf amplificador de la selecció ja que aquesta distribució té una probabilitat de fixació associada major que la probabilitat

de fixació estàndard de Moran,

$$\rho_{estrella} = \frac{1 - \frac{1}{r^2}}{1 - \frac{1}{r^{2n}}} > \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^n}} = \rho_{canonica}.$$

Els grafs amplificadors tenen un paràmetre associat. Diem que un graf amplificador té paràmetre  $K$  quan un mutant al graf amplificador es comportaria de la mateixa manera que ho faria un mutant amb  $r = r^K$  a un graf “normal” (ni supressor ni amplificador). En particular, l'estrella és un amplificador de paràmetre 2. És per això que la probabilitat de fixació és la de Moran estàndard substituint  $r$  per  $r^2$ .

## Capítol 6

# Temps d'absorció

En aquesta secció demostrarem que el procés de Moran sobre un graf  $G = (V, E)$  connex no dirigit amb  $|V| = n$  acaba a un dels dos estats d'absorció en un nombre polinòmic de passos. Tal com s'explica a la secció 3 de [2], veurem que l'esperança d'anar de l'estat  $i$  a  $i + 1$  és més gran que una certa constant. En aquest treball ho demostrarem per a grafs evolutius amb  $r < 1$ . Així doncs, observarem que el procés no s'estanca fins que arriba a un dels estats absorbents.

Començarem definint la funció potencial

$$\phi(S) = \sum_{x \in S} \frac{1}{\deg(x)}$$

per a qualsevol estat  $S \subseteq V(G)$  i denotarem  $\phi(V(G))$  per  $\phi(G)$ . Observem que  $1 < \phi(G) < n$  i que  $\phi(\{x\}) = \frac{1}{\deg(x)} \leq 1$  per qualsevol vèrtex  $x \in V$ .

Primer, demostrarem que l'esperança del potencial augmenta estrictament quan  $r > 1$  i que disminueix estrictament quan  $r < 1$ .

**Lema 8.** Sigui  $(X_i)_{i \geq 0}$  un procés de Moran sobre el graf  $G = (V, E)$  i  $S$  un conjunt tal que  $\emptyset \subset S \subset V$ . Si  $r \geq 1$ , tenim

$$\mathbb{E}[\phi(X_{i+1}) - \phi(X_i) \mid X_i = S] \geq \left(1 - \frac{1}{r}\right) \cdot \frac{1}{n^3},$$

donant-se la igualtat si i només si  $r = 1$ . Si  $r < 1$ ,

$$\mathbb{E}[\phi(X_{i+1}) - \phi(X_i) \mid X_i = S] < \frac{r-1}{n^3}.$$

*Demostració.* Denotem per  $W(S) = n + (r-1)|S|$  l'adaptació genètica total de la població. Per  $\emptyset \subset S \subset V$  i per a qualsevol valor de  $r$ , tenim

$$\begin{aligned} \mathbb{E}[\phi(X_{i+1}) - \phi(X_i) \mid X_i = S] &= \\ &= \frac{1}{W(S)} \sum_{\substack{xy \in E \\ x \in S, y \in \bar{S}}} \left( r \cdot \frac{\phi(S+y) - \phi(S)}{\deg(x)} + \frac{\phi(S-x) - \phi(S)}{\deg(y)} \right) \\ &= \frac{1}{W(S)} \sum_{\substack{xy \in E \\ x \in S, y \in \bar{S}}} \left( r \cdot \frac{1}{\deg(y)} \cdot \frac{1}{\deg(x)} - \frac{1}{\deg(x)} \cdot \frac{1}{\deg(y)} \right) \\ &= \frac{r-1}{W(S)} \sum_{\substack{xy \in E \\ x \in S, y \in \bar{S}}} \frac{1}{\deg(x) \deg(y)}. \end{aligned}$$

Això és clarament 0 si  $r = 1$ . Altrament, podem minimitzar la suma en valor absolut observant que ha d'haver, com a mínim, una aresta entre  $S$  i  $\bar{S}$  i, els seus vèrtexs tindran com a màxim grau  $(n-1) < n$ . A més, tenim que l'estat amb  $W(S)$  més gran si  $r > 1$  és el de fixació i, per altra banda, si  $r < 1$  és el d'extinció.

Així doncs, per a  $r > 1$  tenim

$$\mathbb{E}[\phi(X_{i+1}) - \phi(X_i) \mid X_i = S] > \frac{r-1}{rn} \cdot \frac{1}{n^2} = \left(1 - \frac{1}{r}\right) \frac{1}{n^3}$$

i, si  $r < 1$  tenim

$$\mathbb{E}[\phi(X_{i+1}) - \phi(X_i) \mid X_i = S] < (r-1) \frac{1}{n^3}.$$

□

A continuació, enunciem un teorema que dóna una fita superior pel temps esperat d'absorció. Els científics He i Yao, van donar una demostració per aquest, tal com podem veure a [5]. Aquesta prova utilitza coneixements de



martingales que s'escapen del nostre camp d'estudi. És per això que ens limitarem a l'enunciació del teorema.

Aquest resultat, ens serà d'utilitat per poder afitar el temps d'absorció associat al procés de Moran per  $r < 1$ .

**Teorema 9.** *Sigui  $(Y_i)_{i \geq 0}$  una cadena de Márkov amb espai d'estats  $\Omega$ , on  $Y_0$  és escollit d'un conjunt  $I \subseteq \Omega$ . Si existeixen constants  $k_1, k_2 > 0$  i una funció no negativa  $\psi : \Omega \rightarrow \mathbb{R}$  tal que*

- $\psi(S) = 0$  per algun  $S \in \Omega$ ,
- $\psi(S) \leq k_1$  per tot  $S \in I$  i
- $\mathbb{E}[\psi(Y_i) - \psi(Y_{i+1}) \mid Y_i = S] \geq k_2 \forall i \geq 0$  i tot  $S$  tal que  $\psi(S) > 0$ ,

aleshores  $\mathbb{E}[\tau] \leq \frac{k_1}{k_2}$ , on  $\tau = \min\{i : \psi(Y_i) = 0\}$ .

Ara doncs, com ja hem dit, donarem una fita superior pel temps d'absorció esperat del procés de Moran.

**Teorema 10.** *Sigui  $G = (V, E)$  un graf de grau  $n$ . Per  $r < 1$  i qualsevol  $S \subseteq V$ , el temps d'absorció  $\tau$  pel procés de Moran per  $G$  satisfà*

$$\mathbb{E}[\tau \mid X_0 = S] \leq \frac{1}{1-r} n^3 \phi(S).$$

*Demostració.* Sigui  $(Y_i)_{i \geq 0}$  el procés de Moran sobre  $G$  exceptuant que, si s'arriba a l'estat de fixació, s'introdueix un nou individu no mutant en el lloc d'un vèrtex escollit aleatòriament. Així, de l'estat  $V$  anem a l'estat  $V - x$  on  $x \in V$  és escollit de forma aleatòria. Definim  $\tau' = \min\{i : Y_i = \emptyset\}$  pel temps d'absorció d'aquest nou procés. Clarament  $\mathbb{E}[\tau \mid X_0 = S] \leq \mathbb{E}[\tau' \mid Y_0 = S]$ .

La funció  $\phi$  compleix les hipòtesis de  $\psi$  del teorema 9 on  $k_1 = \phi(S)$  i  $k_2 = (1-r)n^{-3}$ . Comprovem que també se satisfan les condicions per  $\phi$ :

- $\phi(S) = 0$  per  $S = \emptyset \in \Omega$ .
- $\phi(S) \leq k_1 = \phi(S)$  per tot  $S \subseteq V$ .
- $\mathbb{E}[\phi(Y_i) - \phi(Y_{i+1}) \mid Y_i = S] \geq k_2 = (1-r)\frac{1}{n^3}$ . Cert pel lema 8 quan  $r < 1$ .

Així doncs, podem aplicar el teorema 9. Fent-ho obtenim

$$\mathbb{E}[\phi(Y_i) - \phi(Y_{i+1}) \mid Y_i = V] = \frac{1}{n} \sum_{x \in V} \frac{1}{\deg x} > \frac{1}{n} > k_2.$$

Aleshores,

$$\mathbb{E}[\tau \mid X_0 = S] \leq \mathbb{E}[\tau' \mid Y_0 = S] \leq \frac{1}{1-r} n^3 \phi(S).$$

□

## Capítol 7

# Problemes oberts

Són molts els grafs evolutius pels quals no s'ha trobat un valor asimptòticament exacte de la probabilitat ni del temps de fixació associats. En aquest treball hem vist que podem obtenir aquests valors per a grafs evolutius regulars i per a grafs amb pocs vèrtexs. Però, hi ha molts altres grafs pels quals el càlcul és massa complex i, de moment, no es coneix ni la probabilitat de fixació ni el temps d'absorció corresponent de forma teòrica. En canvi, sí que s'han trobat aproximacions d'aquests valors experimentalment mitjançant simulacions amb computadores. En aquesta secció, presentarem dos problemes oberts: la doble estrella i el camí. Ambdós grafs són no dirigits. Quan parlem de cadascun d'aquests, remarcarem la importància d'aquest fet.

### 7.1 Doble estrella

Aquest graf evolutiu consisteix en dues estrelles unides per una aresta, tal com mostra la Figura 7.1. D'aquest graf no es coneix la probabilitat de fixació ni el temps esperat d'absorció. De moment, l'únic que sabem és

- $\rho \leq 1 - \frac{1}{n+r}$  si  $r > 0$  pel lema 4.
- $\mathbb{E}[\tau \mid X_0 = S] \leq \frac{1}{1-r} n^3 \phi(S)$  si  $r < 1$  pel teorema 10.

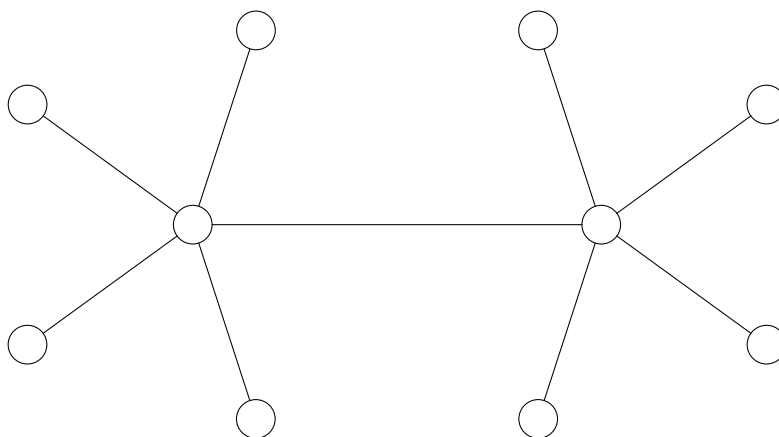


Figura 7.1: Doble estrella

## 7.2 Camí

Per altra banda, altre problema obert és el camí, que té forma de recta no dirigida, tal com podem veure a la Figura 7.2. La probabilitat de fixació associada a aquest graf ha estat estudiada per diversos científics. Esmentarem a M. Broom i J. Rychrář ja que el mateix article que hem utilitzat com a font d'informació per a diversos punts d'aquest treball, conté també el càlcul no exacte de la probabilitat de fixació associada al graf amb forma de camí. El lector pot conèixer més sobre el raonament que van seguir consultant la secció 6 de [1].

Figura 7.2: Camí no dirigit de  $n$  vèrtexs.

Ara, observarem com d'important és que el graf sigui no dirigit en aquest cas per a que continuï sent un problema obert. Estudiem doncs, el graf evolutiu de la Figura 7.3 (cas anàleg al de la Figura 7.2 però amb arestes dirigides unidireccionalment entre els vèrtexs). Notem que el raonament sobre aquest graf és elemental. Això és conseqüència de què cada vèrtex és influenciat, exclusivament, per l'anterior. És a dir, el vèrtex  $n$  només pot ser influenciat

pel vèrtex  $n - 1$ , l' $n - 1$  només pot ser influenciat per l' $n - 2$ , ... i així successivament fins arribar a que el vèrtex 2 només pot ser influenciat pel vèrtex 1. D'aquesta manera, tenim que l'espècie que acabarà dominant el graf evolutiu serà la del vèrtex 1: si el vèrtex 1 representa un individu mutant, la resta de vèrtexs s'acabaran convertint en mutants ja que cap d'ells és capaç de transformar el vèrtex 1 en un no mutant. El mateix succeeix si 1 representa a un individu no mutant. És per això que la probabilitat de fixació és  $\rho = \frac{1}{n}$  ja que és la probabilitat de què el vèrtex 1 sigui mutant a temps 0.

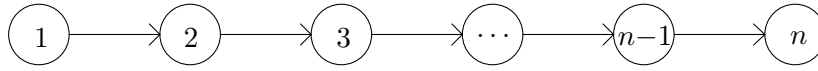


Figura 7.3: Camí dirigit de  $n$  vèrtexs.

Acabem de veure que donat un graf del qual no es coneix el valor asimptòticament exacte per a la probabilitat de fixació, si afegim direccions a les arestes, transformem el càlcul de la probabilitat de fixació en un fet elemental. A continuació, presentarem un altre graf interessant de ser comentat.

Aquest és el graf evolutiu amb forma d'estrella, anàleg al de la Figura 5 però, amb arestes dirigides, tal com mostra la Figura 7.4.

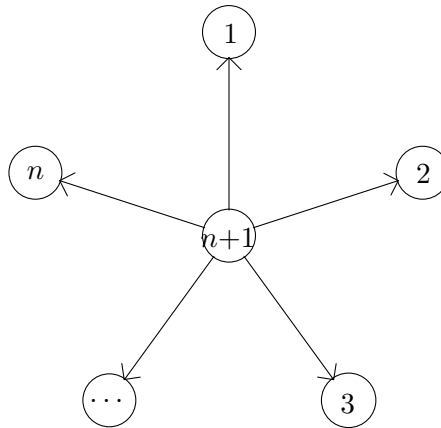


Figura 7.4: Estrella dirigida

Aquest graf té probabilitat de fixació  $\rho = \frac{1}{n+1}$ . El raonament és anàleg al del camí dirigit: la probabilitat de què tots els individus es converteixin

en mutants és igual a la probabilitat de què el vèrtex  $n + 1$  estigui infectat a temps 0.

Malgrat que hem vist dos exemples de grafs amb càlculs molts senzills, són molts els grafs evolutius pels quals no coneixem fites ajustades de la probabilitat de fixació ni del temps d'absorció. Lieberman, com podem veure a [6], ha estudiat alguns dels grafs que ara hem esmentat. Però no ha estat l'únic. Diversos científics han investigat per donar resposta als problemes que hem presentat.

Fixem-nos, però, que l'interès en aquest camp va començar fa uns 60 anys. És a dir, aquesta és una àrea d'estudi relativament nova. Així doncs, tots els resultats que hem presentat en aquest treball han estat obtinguts per científics pioners que han volgut investigar sobre l'evolució i les matemàtiques que hi ha darrere.

# Bibliografia

- [1] M. Broom and J. Rychrář. An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proceedings of the Royal Society A*, 464(2098):2609–2627, 2008.
- [2] Josep Díaz, Leslie Ann Goldberg, George B. Mertzios, David Richerby, Maria J. Serna, and Paul G. Spirakis. Approximating fixation probabilities in the generalized moran process. *Algorithmica*, 69(1):78–91, 2014.
- [3] David Easley and Jon Kleinberg. *Networks, Crowds and Markets. Reasoning about a highly connected world*. Cambridge University Press, 2010.
- [4] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes (3rd Edition)*. Oxford University Press, 2001.
- [5] Yao X. He, J. Drift analysis and average time complexity of evolutionary algorithms. *Artif. Intell.*, 127:57–85, 2001.
- [6] Erez Lieberman, Christoph Hauert, and Martin A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, 2005.
- [7] Martin A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard University Press, 2006.